

University of Groningen

## A taxonomy of care for children and adolescents with behavioural and emotional problems

Evenboer, Klaasje Elisabeth

**IMPORTANT NOTE:** You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

*Document Version*

Publisher's PDF, also known as Version of record

*Publication date:*

2015

[Link to publication in University of Groningen/UMCG research database](#)

*Citation for published version (APA):*

Evenboer, K. E. (2015). *A taxonomy of care for children and adolescents with behavioural and emotional problems: development and application*. [Thesis fully internal (DIV), University of Groningen]. [S.n.].

### Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

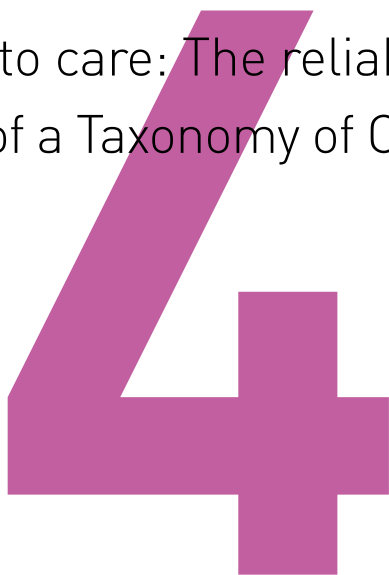
The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

### Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

## Looking into care: The reliability and feasibility of a Taxonomy of Care for Youth



Accepted for publication as: Evenboer, K.E., Huyghen, A.M.N., Tuinstra, J., Reijneveld, S.A., & Knorth, E.J. (2014). Looking into care: The reliability and feasibility of a Taxonomy of Care for Youth. *Journal of Social Work*.

## Abstract

**Summary** – Classifying the care provided to children with emotional and behavioural problems can provide empirical insights into the relationship between child characteristics, the care offered, and outcomes after leaving care. The Taxonomy of Care for Youth (TOCFY) has recently been shown to validly classify this care in six domains covering all aspects of care. The aim of this study was to assess the inter-rater reliability and feasibility of TOCFY. Two raters independently classified the care provided to 200 children (50 per organisation) from organisations in primary health care (PHC), child and youth care (CYC), and mental health care (MHC; two organisations), based on their care records. We assessed inter-rater reliability and the degree to which TOCFY categories could be applied anyhow, i.e. its feasibility.

**Findings** – Mean agreement was 89.8% between raters; excluding the cases scored as “unknown”, the mean agreement was 82.2%. TOCFY -categories could be applied for over 90% regarding each of the six domains.

**Applications** – TOCFY is a valid, reliable and feasible instrument to classify care within different types of care organisations. Given these promising findings, application and further evaluation of TOCFY is recommended.

### KEYWORDS

Social work, reliability, taxonomy, mental health, child and family welfare, social work practice

### ABBREVIATIONS

TOCFY (Taxonomy of Care for Youth), PHC (Primary Health Care), CYC (Child and Youth Care), MHC (Mental Health Care)

## Introduction

The assessment of care provided to emotionally and behaviourally disturbed children and their families can help to determine the relationship between problem behaviour, the care provided, and outcomes (Farmer, Burns, Phillips, Angold, & Costello, 2003; Maschi, Swalbe, Morgen, Gibson, & Violette, 2008). This may in turn support the examination of the quality and effectiveness of the care offered to children and their parents (Courtney & Thoburn, 2009; Fein, 2003; Horwitz et al., 2001; Lee & Barth, 2011; Libby, Coen, Price, Silverman, & Orton, 2005; Ryan & Schuerman, 2004; Sinclair, 2010). To make conclusions about the effectiveness of interventions, more information is needed on the contents of these interventions.

No consensus has been reached about the best measurement of the contents of the care offered (Bjorbekkmo, Myklebust, Olstad, Molvik, Nymann, & Sørgaard, 2009; Lloyd-Evans, Johnson, & Slade, 2007). In the field of child, youth and family care several taxonomic instruments have been developed that aim to classify the care offered to children with behavioural or emotional problems and their families (Evenboer, Huyghen, Tuinstra, Reijneveld & Knorth, 2012a). However, none of these instruments is fully capable to measure the most salient aspects of the care process, i.e. the intervention contents, the intervention recipient, the expertise of the professional, the duration, the intensity and the environment in which the intervention took place (Evenboer, Huyghen, Tuinstra, Knorth & Reijneveld, 2012b).

Besides, information generally lacks about the psychometric qualities of the taxonomic instruments that are already available (Evenboer et al., 2012a). Only Horwitz et al. (2001), Abraham and Michie (2008), and Michie, Hyder, Walia and West (2011) have reported on the reliability of their instruments. This lack of information has implications for the delivery and evaluation of treatment programs, as even if taxonomic information is available, it is of uncertain reliability and validity. In summary, there is an urgent need for care taxonomic instruments in the field of child care and treatment with documented reliability and validity (Horwitz et al., 2001; Michie et al., 2011).

A second criticism on the scarcely available instruments concerns the feasibility of their use. Even if reliable and valid, an instrument can only add to daily practice if it can be easily applied, otherwise hindering the systematic recording of treatment characteristics (Van Yperen et al., 1999). This is in particular problematic for taxonomic systems that concern care for youth with behavioural and emotional problems, given the wide range of care providers involved. Care providers vary from primary care centres to specialised services like youth and social care, and youth mental health care. Each of these types of organisations employ their own terminology to describe clients, care and treatment programmes (Beenackers, 1995; 1999; Dubbeldam et al., 1994; Knorth et al., 2003; Prince, 1996; Van den Bergh, 1992). The variation is probably due both to characteristics of the population as served like type and severity of the problems experienced, and to different theoretical and practical foundations, e.g. coming from a healthcare tradition like mental health care, or from a welfare tradition like youth and social care (Kemp, Whittaker, & Tracy, 1997).

This study focuses on the Taxonomy of Care for Youth (TOCFY), a taxonomic system that has been developed in the Netherlands to classify care and treatment across a wide range of organisations that provide care to youth with behavioural and emotional problems (Evenboer et al., 2012b). TOCFY consists of six domains, i.e. the content, judicial context, duration, intensity, recipients, and the expertise of the professional (Evenboer et al., 2012b). TOCFY was developed by following an empirical procedure, consisting of literature review, expert interviews and consensus meetings of professionals, within the context of the Collaborative Centre on Care for Children and Youth with behavioural and emotional problems (C4Youth). This centre—a collaboration between university departments, care providers, and local authorities in the northern part of the Netherlands—covers the entire range of organisations in the area providing services for the youth concerned.

Within the Netherlands there are roughly three different types of care for children with behavioural and emotional problems, namely primary health care (PHC), child and youth care (CYC), and mental health care (MHC). The first type of care covers relatively mild psychosocial problems; the other two types offer care for more severe problems whereby historically CYC is focused somewhat more on (externalizing) behaviour and family problems whereas MHC is more directed to treatment of emotional and psychiatric problems of the child. In reality there is also quite some overlap (Van Eijk, Verhage, Noordik, Reijneveld, & Knorth, 2013). The way of financing these types of care is heterogeneous, e.g. partially insurance-based and partially government-financed, which causes differences in the entry of children into the care system, i.e. the organisation via which the child enters the system.

The empirical procedure led to a taxonomy with high face validity (Evenboer et al., 2012b), but with not yet documented reliability and feasibility. Therefore, the aim of this study was to assess (1) the inter-rater reliability and (2) the feasibility of TOCFY, and potential differences in outcomes between organisations.

## Methods

### **SAMPLE**

We collected data on a stratified random sample of children and adolescents (N=200; 4-18 years), receiving care at one of the four participating care organisations in the C4Youth study (Table 4.1). These organisations were (number of clients participating in the C4Youth study in the period June 2011-June 2012 between brackets): (a) primary health care (PHC: 339), (b) child and youth care (CYC: 150), and (c) mental health care (MHC-A: 200; MHC-B: 429). From each of these four care organisations, 50 care records were randomly chosen of the group of children who participated in the C4Youth study by using a random number generator. This selection was done by the first author with the help of the administrative staff within each of the mentioned care organisations. For each child that participated in the C4Youth study, professionals filled in information about the care offered. In total it concerns 114 professionals (male N=16, female N=98; age M=40.0, SD=10.8; years of experience in care M=12.7,

SD=9.5; years of experience within youth care M=9.9, SD=8.4). Because we used the same sample of 200 children as for assessing inter-rater reliability, some professionals - having two or more children of the sample in their caseloads - had to fill in more than one questionnaire.

**TABLE 4.1** Age and gender of children and adolescents participating in the reliability and feasibility analyses

	4-11 years	12-18 years	Total
Sex			
Male	51 (25.5%)	54 (27.0%)	105 (52.5%)
Female	38 (19.0%)	57 (28.5%)	95 (47.5%)
Total	89 (44.5%)	111 (55.5%)	200 (100%)

#### **TOCFY: DOMAINS AND SCORING CATEGORIES**

TOCFY consists of six domains (Evenboer et al., 2012b). The content of the intervention is the most elaborate domain, describing the interventions offered by the four care organisations. For example, the intervention 'care program for behavioural disorders' of the MHC-B organisation consists of a diagnostic and a treatment module. The treatment module is then further elaborated into ten subcategories like, for example, 'psycho education', 'system therapy', 'medication' and 'cognitive behavioural therapy'. Scoring took place at the categories and subcategories level. The judicial context indicates whether care was offered voluntarily, whether a coercive or compulsory placement was applied, or whether care was offered within the context of a civil or penal placement. The duration of the intervention ranges from less than four weeks of care to a stay of more than a year in care. The domain of intensity of intervention concerns the average number of contacts per day/week/month/year and the average number of minutes per contact. The domain of intervention recipients covers the entire range of people and environments that could be involved in the treatment. The expertise of the professional involved in the treatment is classified in the expertise domain. Each domain contains the category 'other, namely...', for when the existing categories prove unsatisfactory. The number of categories varies per domain, and regarding 'content' also per organisation (Table 4.2).

#### **PROCEDURE**

For assessing the inter-rater reliability of TOCFY, care records of children (N=200) were assessed by two independent assessors (MSc students at the University of Groningen, who were supervised by the first author), using TOCFY. The assessors were instructed beforehand on how to use TOCFY and assess care records in order to limit any differences in procedure. For each intervention, the raters scored per domain one category as provided by TOCFY, except for 'intensity'. For that domain, raters scored two categories: the mean number of contacts and the mean number of minutes per contact. Moreover,

**TABLE 4.2** Number of TOCFY categories per domain (specified by organization for the “content” domain)

Domain	Description	Number of categories
1. Content	What care/intervention is being offered? · PHC* · CYC · MHC-A · MHC-B	7 42 31 11
2. Judicial context	In which judicial context is care being offered?	12
3. Duration	What is the duration of the care being offered?	9
4. Intensity Sub-domains	What is the intensity of the care being offered? <i>Mean number of contacts</i> <i>Mean number of minutes per contact</i>	5 <i>Exact number</i> <i>Exact number</i>
5. Recipients	Who are receiving care?	13
6. Expertise	What is the expertise of the main professionals involved?	32

for the domains ‘content’ and ‘judicial context’ they could sometimes rate sub-categories, i.e. more detailed categories, as part of a more general category. For these domains, the inter-rater reliability was determined at the level of the general categories only, to make outcomes comparable to those regarding the other domains. Data were collected between February and June 2012.

Feasibility was assessed by measuring how well professionals were able to apply the six TOCFY domains on care as provided. Therefore, the child’s main therapist was asked to apply TOCFY to all care he or she provided during the first three months of the treatment. This concerned the same 200 real-life cases as used to assess the inter-rater reliability.

## DATA ANALYSIS

First, inter-rater reliability was assessed as percentages of agreement between the two raters for the six domains (Flack, Afifi, Lachenbruch, & Schouten, 1988; Fleiss, 1981), overall and per organisation. The percentage of agreement between raters was the most appropriate measure for inter-rater reliability (Grayson & Rust, 2001). We did not use Cohen’s Kappa because many TOCFY domains contained too many categories for this (Cohen, 1960). In some cases, no information was available on specific domains and the raters scored the category ‘Unknown or not otherwise specified’. Because this might affect measures of agreement, we assessed agreement with inclusion and with exclusion of these missing values.

Second, we assessed inter-rater reliability for numerical variables by using Bland and Altman (1986) plots. This concerned the domain ‘intensity’, measuring the mean number of contacts and the mean number of minutes per contact. Bland and Altman plots show whether the scores of both raters regarding the number of contacts and the number of minutes per contact are within the ‘limits of agreement’ ( $\text{Mean} \pm 2\text{SD}$ ). If the scores of both raters were within these limits, the differences between the scores were considered acceptable. Bland and Altman plots could only be constructed for both MHCs; for the PHC and the CYC organisations information regarding intensity of treatment was often missing.

Third, we assessed feasibility by determining the percentage of cases in which professionals were able to apply TOCFY per domain, overall and per organisation.

## Results

### INTER-RATER RELIABILITY

Overall, mean agreement was 89.8% for the six TOCFY domains (Table 4.3), ranging from 80.5% for ‘recipients’ to 99.0% for ‘judicial context.’ After excluding the missing values, the mean agreement for the six domains was 82.2%, ranging from 69.5% for ‘duration’ to 98.7% for ‘judicial context.’ Exclusion of missing values was not meaningful for the PHC organisation regarding the domains ‘judicial context’, ‘intensity’, and ‘recipients’, and for the CYC organisation regarding ‘duration’ and ‘expertise’; for these domains the number of missings exceeded the number of valid scores. The category ‘other, namely...’ was not frequently used by the professionals; only in 4.5% of the cases this category was used.

In addition to the percentages of agreement, we used Bland and Altman (1986) plots to explore

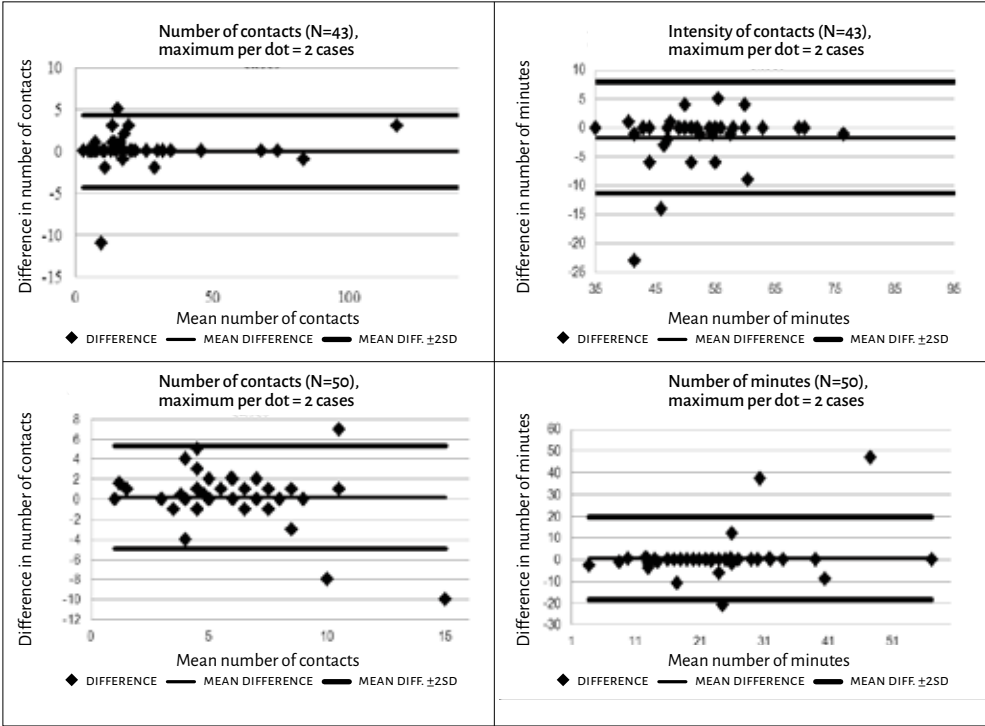
**TABLE 4.3** Inter-rater reliability for the TOCFY domains: percentages of agreement for all organizations combined and for each care organization separately

Domain	Intervention content	Judicial context	Duration	Intensity	Recipients	Expertise	Across all domains*
N (including missing values)	200	200	200	200	200	200	1200
Percentage of agreement All organizations combined	88.0%	99.0%	89.0%	98.5%	80.5%	84.0%	89.8%
PHC**	94.0%	100%	76.0%	100%	100%	90.0%	93.3% (300)
CYC**	90.0%	96.0%	100%	96.0%	72.0%	100%	92.3% (300)
MHC-A**	80.0%	100%	98.0%	98.0%	92.0%	78.0%	91.0% (300)
MHC-B**	88.0%	100%	82.0%	100%	58.0%	68.0%	82.7% (300)
N (excluding missing values)	200	150	62	116	142	114	784
Percentage of agreement All organizations combined	88.0%	98.7%	69.5%	93.5%	71.3%	71.9%	82.2%
PHC	94.0%	NA***	67.7%	NA***	NA***	90.0%	83.9% (134)
CYC	90.0%	96.0%	NA***	82.4%****	72.0%	NA***	85.1% (167)
MHC-A	80.0%	100%	80.0%	98.0%	92.0%	78.7%	88.1% (252)
MHC-B	88.0%	100%	60.9%	100%	50.0%	47.1%	74.3% (231)

\* The total number across all domains is the sum of the number of care records for each domain of TOCFY separately. \*\* PHC = Primary Health Care, CYC = Child and Youth Care, MHC = Mental Health Care. \*\*\* NA (Not Applicable) refers to when both raters recorded “unknown” for each case in a specific domain. In these instances a percentage of agreement cannot be determined. \*\*\*\* For the CYC an agreement percentage without missing values could be measured for the intensity domain (N=6).



differences in the numerical scores of the raters in the ‘intensity’ domain (Figure 4.1). The agreement between raters was acceptable (Mean  $\pm$  2SD), except for some outliers. The mean number of contacts was 25.9 (SD=2.2) for MHC-A, and 5.5 (SD=2.6) for MHC-B. The mean number of minutes per contact was 52.3 (SD=4.8) for MHC-A, and 22.8 (SD=9.5) for MHC-B.



**FIGURE 4.1** Plot of differences between two raters scoring the number of contacts and number of minutes in MHC-A (N=43) and MHC-B (N=50)

### FEASIBILITY

Analyzing the information that had to be filled in concerning TOCFY resulted in scores above 90% for all the six domains. The higher the percentages, the better the professionals were able to use TOCFY for mapping information about the care offered. The category ‘other, namely...’, has not been used frequently by the professionals. Professionals of MHC-B were most likely to complete the information concerning the domains of TOCFY and professionals of the MHC-A were least likely to complete TOCFY entirely.

## Discussion

Aim of this study was to assess the inter-rater reliability and feasibility of TOCFY, overall and per type of organisation. The results showed good inter-rater reliability for TOCFY, with an overall inter-rater agreement of 89.8% (including missing values) and 82.2% (excluding missing values). Categories of the six domains of TOCFY could be applied for over 90%, representing a good feasibility.

Inter-rater reliability was mostly good, especially given the complex nature of the taxonomic system, but was somewhat lower for the domains 'duration' and 'expertise' when excluding the missing values. For the 'duration' domain, an explanation could be that some interventions were still provided to the client at the moment of the study, making it impossible to measure the realized duration of the care. For the 'expertise' domain an explanation could be that frequently more than one professional was involved during the treatment trajectory. This caused some minor differences in the categories scored by the raters.

The previous studies that assessed inter-rater reliability of taxonomic instruments (Abraham & Michie, 2008; Michie et al., 2011) found comparable agreement, ranging from 85% up to 95%, but these concerned taxonomies for a different type of care, i.e. behavioural and lifestyle change. Besides, these taxonomies (Abraham & Michie, 2008; Michie, 2011) were applied to intervention manuals and not to actually provided care. Therefore our findings on reliability should be confirmed by other studies on real care settings.

A taxonomic instrument is of no use for practice if it is not feasible for professionals (Van Yperen et al., 1999). Fortunately TOCFY proved to be feasible in all organisations, albeit to a somewhat varying degree. A rather likely explanation for this variation would be variation in the type of care offered. However, interestingly, the largest differences concerned two organisations that offered a similar type of care, i.e. child mental health care. The highest feasibility was measured for MHC-B and the lowest for MHC-A. An alternative explanation for the relatively low percentage in MHC-A might be that some professionals could not complete the questionnaire because of practical reasons such as time pressure.

To optimise the feasibility of TOCFY, professionals of the participating care organisations were closely involved in the development procedure (Evenboer et al., 2012b). This resulted for example in the use of organisation specific terminologies to describe the care offered for each of the four care organisations. By doing so, the manageability of the instrument in daily practice was optimized. The use of TOCFY may not only lead to more information about the care that is offered, but also to using more standardized terminologies and categories for describing care. As a consequence, the communication of professionals within or between care organizations can become more transparent. This can in turn improve the efficiency of the whole care trajectory. It implies too, however, that TOCFY should be kept up to date to the current care offer, which is a continuous, but necessary, process for a clinically applied taxonomy.

## **STRENGTHS AND LIMITATIONS**

A major strength of this study is the use of care records from a range of organisations providing care for children and adolescents with behavioural and emotional problems. This allowed us to gain an overview of TOCFY's reliability and feasibility in different types of care organisations. Another strength is that the same sample of children (and parents) who received real care was used to measure inter-rater reliability and feasibility. A third strength is that we used the percentage of agreement to measure inter-rater reliability, which is the most appropriate method for measuring the inter-rater reliability of an instrument containing multiple categories and sub-categories (Grayson & Rust, 2001).

A limitation of the study is we had to rely on the information as registered per case in the care records. Ideally, additional information would have been gathered as a way of 'methodological triangulation' (Denzin, 2006; Rothbauer, 2008), for instance by real-time observation in care organisations and interviews with professionals providing care to these children and their families. Due to time limits and the amount of cases that should have been followed (N=200) this was not a realistic option. Another limitation of the present study is that only the inter-rater reliability was assessed. Other measures of reliability might be used in addition. The test-retest reliability, for example, could be assessed in a future study by applying TOCFY to a new set of care records, thereby asking the professionals involved to participate in two rounds of assessment within a given period of time.

## **IMPLICATIONS**

Our findings have implications for practice and further research into the care offered to emotionally and behaviourally disturbed children and their families. TOCFY seems to be a feasible, valid, and reliable instrument, meaning that professionals could use TOCFY in daily practice to gather and monitor information about the care offered to children and their families. The fact that TOCFY is a valid and manageable instrument may reflect the close involvement of professionals in the empirical development procedure (Evenboer et al., 2012b). As such, TOCFY is ready to be used in other care settings than included in our study.

TOCFY is currently being used in the C4Youth study (Knorth, Reijneveld, Van Eijk, Noordik, & Tuinstra, 2011), enabling us to gather more information about the care that is offered to children and their families within primary health care, child and youth care, and mental health care. Care organisations themselves could also make use of the domains and categories of TOCFY for a more systematically recording of information in care records about the care actually offered to children and their families. In this way, it is possible to determine whether children with more severe problems receive other treatments, for example more intensive, than children who suffer from milder problems.

In the future applying TOCFY may thus support the examination of problem characteristics, care offered, and outcomes. In this way, a part of the 'black box' of care can be opened. This in turn can help to further optimize the care offered to these children and their families (Ezell et al., 2011; Fein, 2003; Sinclair, 2010). Moreover, feasibility, validity and reliability of TOCFY in other settings deserve additional study.

## **CONCLUSION**

We conclude that TOCFY is a valid, reliable and feasible instrument for assessing the most salient characteristics of care and treatment within different types of care organisations. We found fairly good psychometric qualities, but these findings should be replicated in other populations and countries.

## References

- Abraham, C., & Michie, S. (2008). A taxonomy of behaviour change techniques used in interventions. *Health Psychology, 27*(3), 379-387.
- Beenackers, A. A. J. M. (1995). A closer look at patient records in mental health care. *Maandblad Geestelijke Volksgezondheid, 50*(7/8), 820-821 (in Dutch).
- Beenackers, A. A. J. M. (1999). Improvement of patient records in mental health care: from a few notes to regular letters. *Maandblad Geestelijke Volksgezondheid, 54*(1), 3-10 (in Dutch).
- Bjorkbekkmo, S., Myklebust, L. H., Olstad, R., Molvik, S., Nymann, A., & Sørgaard, K. (2009). Decentralization matters. Differently organized mental health services relationship to staff competence and treatment practice: the VELO study. *International Journal of Mental Health Systems, 3*(1), 9-19.
- Bland, M. J., & Altman, D. G. (1986). Statistical methods for assessing agreement between two methods of clinical measurement. *The Lancet, 327*(8476), 307-310.
- Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement, 20*(1), 37-46.
- Courtney, M., & Thoburn, J. (eds.) (2009). *Children in State Care*. Farnham, UK / Burlington, VT: Ashgate.
- Denzin, N. (2006). *Sociological methods: A sourcebook*. Piscataway, NJ: Aldine Transaction Publishers (5th edition).
- Dubbeldam, J. W. E., Stapper, E. A. A., & Knorth, E. J. (1994). Referral reports in residential child and youth care: A study of the informative quality. *Kind en Adolescent, 15*(1), 46-51 (in Dutch).
- Evenboer, K. E., Huyghen, A. M. N., Tuinstra, J., Reijneveld, S. A., & Knorth, E. J. (2012a). Taxonomic systems in the field of health care, family care, and child and youth care: A systematic overview of the literature. *Children and Youth Services Review, 34*, 2304-2310.
- Evenboer, K. E., Huyghen, A. M. N., Tuinstra, J., Knorth, E. J., & Reijneveld, S. A. (2012b). A taxonomy of care for youth: Results of an empirical development procedure. *Research on Social Work Practice, 22*(6), 637-646.
- Ezell, M., Spath, R., Zeira, A., Canali, C., Fernandez, E., Thoburn, J., & Vecchiato, T. (2011). An international classification system for child welfare programs. *Children and Youth Services Review, 33*(10), 1847-1854.
- Farmer, E. M. Z., Burns, B. J., Phillips, S. D., Angold, A., & Costello, E. J. (2003). Pathways into and through mental health services for children and adolescents. *Psychiatric Services, 54*(1), 60-66.
- Fein, E. (2003). The black box. Accounting for program inputs when assessing outcomes. In A. N. Maluccio, C. Canali, & T. Vecchiato (eds.), *Assessing outcomes in child and family services* (pp. 19-27). New York: Walter de Gruyter, Inc.
- Flack, V. F., Afifi, A. A., Lachenbruch, P. A., & Schouten, H. J. (1988). Sample size determinations for the two rater kappa statistic. *Psychometrika, 53*(3), 321-325.
- Fleiss, J. L. (1981). *Statistical methods for rates and proportions*. New York: John Wiley & Sons, Ltd.
- Grayson, K., & Rust, R. (2001). Interrater reliability. *Journal of Consumer Psychology, 10*(1/2), 71-73.
- Horwitz, S. M., Hoagwood, K., Stiffman, A. R., Summerfeld, T., Weisz, J. R., Costello, E. J., Rost, K., Bean, D. L., Cottler, L., Leaf, P. J., Roper, M., & Norquist, G. (2001). Reliability of the services assessment for children and adolescents. *Psychiatric Services, 52*(8), 1088-1094.
- Kemp, S. P., Whittaker, J. K., & Tracy, E. M. (1997). *Person-Environment Practice. The social ecology of interpersonal helping*. Hawthorne, NY: Aldine De Gruyter.
- Knorth, E. J., Metselaar, J., Josias, H. J., Konijn, C., Noom, M. J., & Van Yperen, T. A. (2003). Indications for treatment in child and youth care: Results from two complementary empirical studies. *International Journal of Child and Family Welfare, 6*(4), 167-184.
- Knorth, E. J., Reijneveld, S. A., Van Eijk, L. M., Noordik, F. W., & Tuinstra, J. (2011). Sticking closely to the social-emotional development of three thousand children inside and outside care. *Kind en Adolescent Praktijk, 10*(4), 192-194 (in Dutch).
- Lee, B. R., & Barth, R. P. (2011). Defining group care programs: An index of reporting standards. *Child and Youth Care Forum, 40*(4), 253-266.
- Libby, A. M., Coen, A. S., Price, D. A., Silverman, K., & Orton, H. D. (2005). Inside the black box: What constitutes a day in a residential treatment center? *International Journal of Social Welfare, 14*(3), 176-183.
- Lloyd-Evans, B., Johnson, S., & Slade, M. (2007). Assessing the content of mental health services: a review of measures. *Social Psychiatry and Psychiatric Epidemiology, 42*(8), 673-682.
- Maschi, T., Hatcher, S. S., Schwalbe, C. S., & Rosato, N. S. (2008). Mapping the social service pathways of youth to and through the juvenile justice system: A comprehensive review. *Children and Youth Services Review, 30*(12), 1376-1385.
- Michie, S., Hyder, N., Walia, A., & West, R. (2011). Development of a taxonomy of behaviour change techniques used in individual behaviour support for smoking cessation. *Addictive Behaviors, 36*(4), 315-319.
- Prince, K. (1996). *Boring records? Communication, speech and writing in social work*. London/Philadelphia: Jessica Kingsley Publishers.
- Rothbauer, P. (2008). Triangulation. In L. Given (ed.), *The SAGE Encyclopedia of qualitative research methods* (pp. 892-894). Thousand Oaks, CA: Sage.

- Ryan, J. P., & Schuerman, J. (2004). Matching family problems with specific family preservation services: A study of service effectiveness. *Children and Youth Services Review*, 26(4), 347-372.
- Sinclair, I. (2010). Inside the black box: What makes for success in care? In E. J. Knorth, M. E. Kalverboer, & J. Knot-Dickscheit (eds.), *Inside out: How interventions in child and family care work. An international source book* (pp. 41-44). Antwerp, Belgium: Garant Publishers.
- Van den Bergh, P. M. (1992). Information about the client. In J. D. van der Ploeg, et al. (eds.), *Vulnerable youth in residential care. Part II: Clients, staff and the system* (pp. 45-57). Leuven: Garant Publishers.
- Van Eijk, L., Verhage, V., Noordik, F. W., Reijneveld, S. A., & Knorth, E. J. (2013). *TakeCare Province of Groningen. A study on the characteristics of children and their parents in relation to the care offered*. Groningen: University of Groningen & University Medical Center Groningen.
- Van Yperen, T. A., Konijn, C., & Ten Berge, I. (1999). *Registration of intended and implemented care. Report of a pilot study on child and youth care services in the province of Gelderland*. Utrecht: NIZW / Trimbos Institute (in Dutch).

